High Performance Computing Facilities for the Next Millennium

Computational Systems

SC99 Tutorial
November 14, 1999

James Craw

Computational Systems, Group Leader

craw@nersc.gov

- **MPP - T3E-900 LC with 696 PEs - UNICOS/mk**

  - **644 Application PEs**

  - **256 MB per PE**

  - **383 GB of Swap Space - 5 partitions, each 5-way striped**

  - **582 GB Checkpoint File System - 5 partitions, striped**

  - **1.5 TB /usr/tmp File System**

  - **7- 25 GB Home File System, DMF managed**

  - **Queuing Systems: NQE/NQS**

- ■ **PVP - One J90 SE Interactive System running UNICOS**
  - ● **32 CPUs**
  - ● **8 GB Memory**
  - ● **Home File Systems NFS Exported to Batch Systems**
    - ♦ **NFS over HPPI $\Rightarrow$ NFS over gigaring**
  - ● **Queuing Systems: NQE/NQS**

- **PVP - Three J90 SV-1 Batch Only Systems running UNICOS**
  - **64 CPUs Total**
  - **8 GB of Memory per System (24 GB total)**
  - **1.0 TB local /usr/tmp**
  - **Queuing Systems: NQE/NQS**

Production history of PVP, MPP, COTS @ NERSC
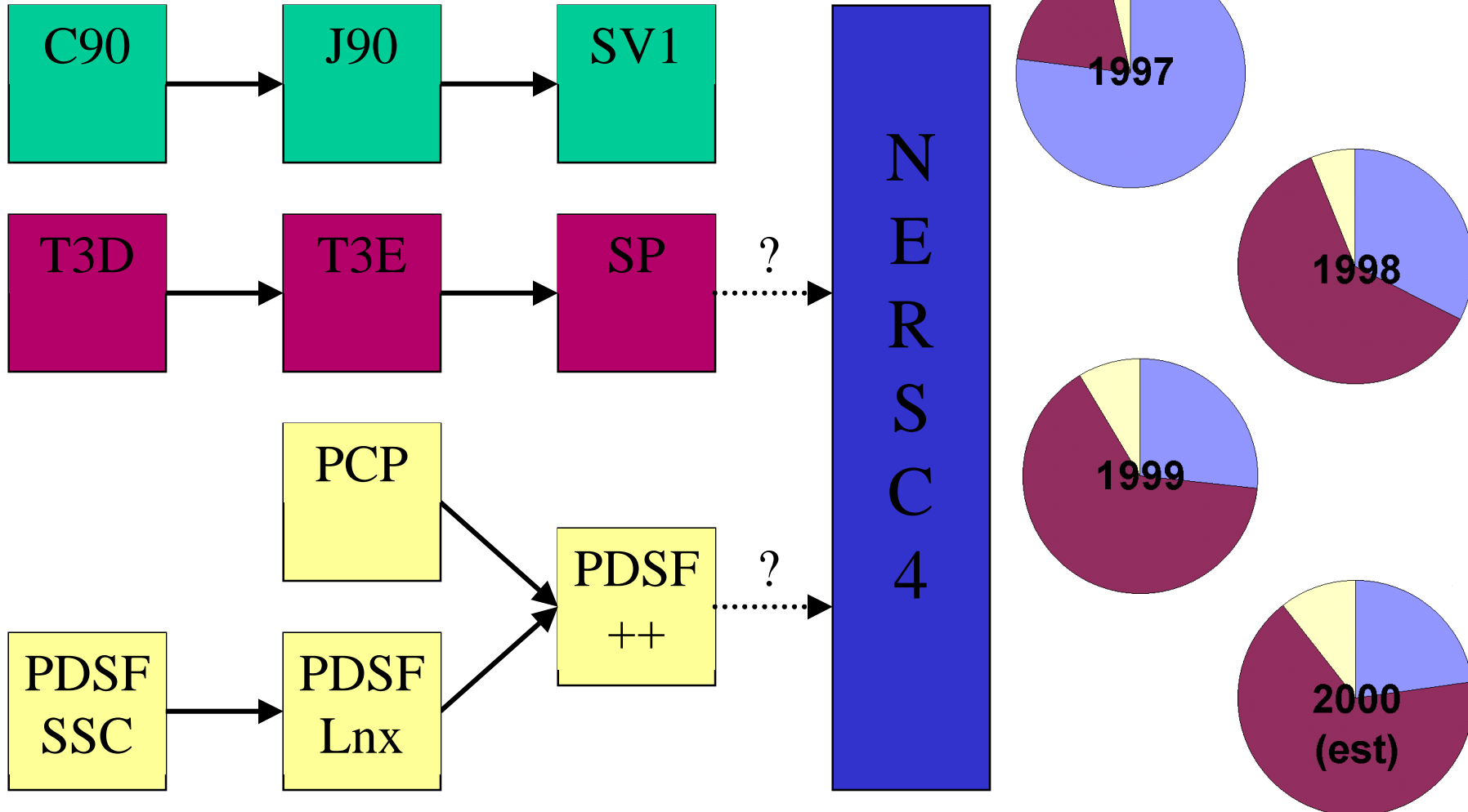
- **MPP**
  - **H/W Evolutions for NERSC**

    T3E-600 LC-136 $\Rightarrow$

    Upgraded to T3E-900 LC-544 $\Rightarrow$

    Added second T3E-600 LC-104 and

    Then Upgraded To T3E-900 LC-153 $\Rightarrow$

    Merged T3E's to make a single T3E-   900 LC-696

  - **S/W Evolutions for NERSC**

    UNICOS/mk (1.0 $\Rightarrow$ 2.0.4.67)

    Since March 1999 there have been 24 archives released. That's 24 in 31 weeks

- **Super Homes (PVP & MPP)**
  - **Free Space Managed Via Hierarchical Storage System**
    - ♦ **HPSS Manages Offline Storage System**
    - ♦ **Data Migration Facility (DMF) Manages Online Storage (User Filesystems)**
      - ➢ **Enabled when filesystem fills to a specified threshold**
      - ➢ **Meta-data portion of user file remains online**
      - ➢ **Data portion of a user file is stored offline via ftp to HPSS**
      - ➢ **Offline data can be recalled explicitly/implicitly by a user**

■ Super Homes (PVP & MPP cont'd)

- ♦ **Recoverability**
- ♦ **Reduces size of nightly backups**
  - ➢ Online files' meta-data and data are backed up
  - ➢ Offline files' meta-data (including DMF retrieval data) only is backed up
- ♦ **User Limits Managed via Filesystem Quotas**
- ♦ **It Works Well!**

- **High System Utilization**
 - **Requires Balanced System**
    - **Stable O/S, Micro Kernal, Single System Image, Source Code**
    - **CPUs**
    - **Adequate Disk space**
      - **Homes (+ DMF)**
      - **/usr/tmp**
      - **System (alternate boot)**
      - **Checkpoint/Restart (3 - 4 x total memory)**
      - **Swap (2 - 3 x total memory)**

● **Balanced System (cont'd)**

   ♦ **I/O bandwidth**

   ♦ **CPUs**

   ♦ **Memory**

   ♦ **Checkpoint/Restart**

   ♦ **Queuing System (NQE/NQS)**

   ♦ **PSCHED for T3E**

      » Gang Scheduling

      » Prime Job

      » Load Balancing (process migration)

- **Minimize idle time in the APP region**

- **Provide fast interactive response while managing the total interactive workload on the system**

- **Provide reasonable and even turnaround across all the batch queues**

- **Encourage users to scale applications to large number of PEs**
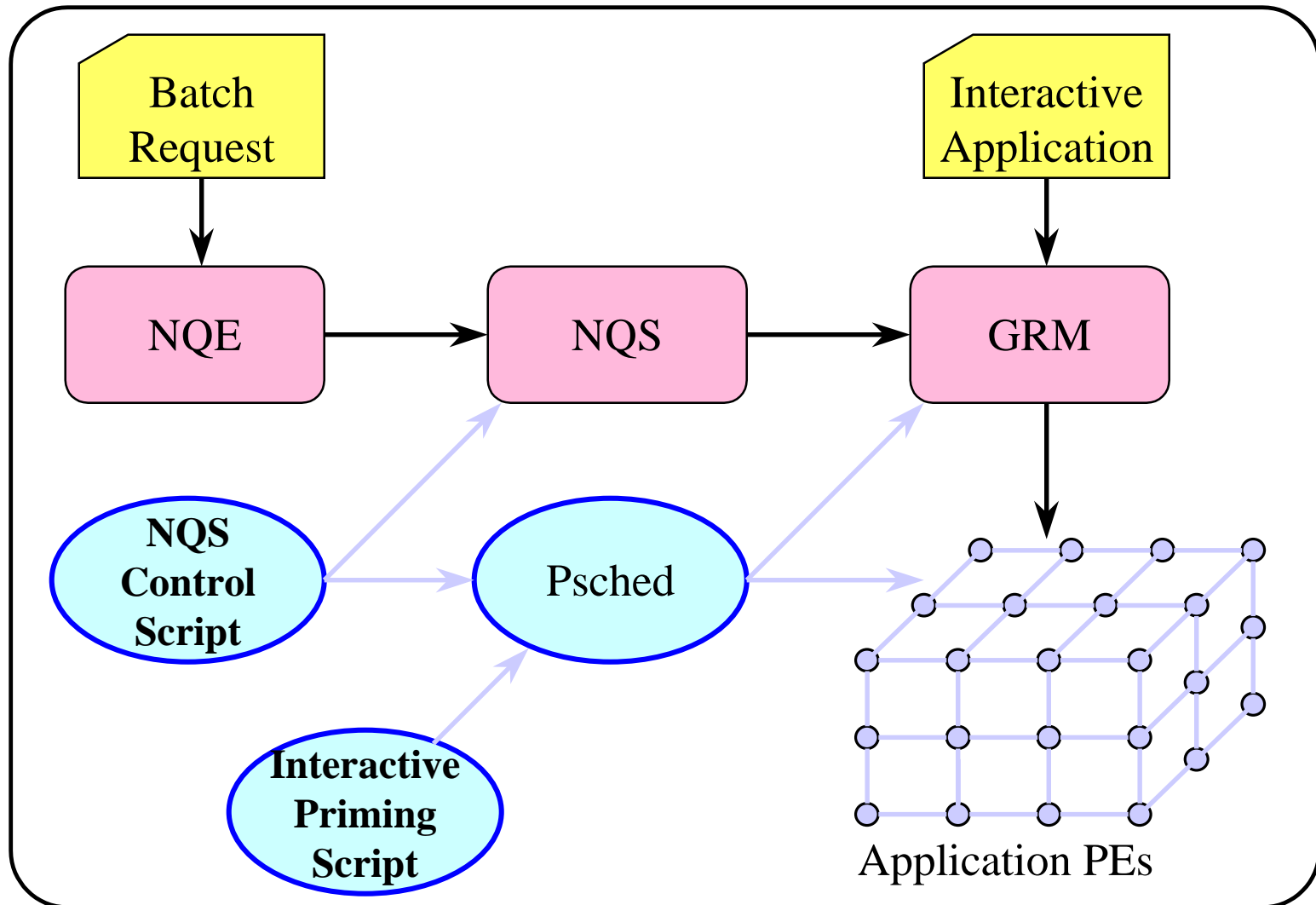
■ **NQE - holding pen for incoming requests**

- ● **Production Queues: LWS limit of 3 jobs per user**
- ● **Debug Queues: LWS limit of 1 job per user**

Queue	P EL im	T ime L im	Pri o ity
Pe 5 1 2	512	4 h r	45
Pe 2 5 6	256	4 h r	30
Pe 1 2 8	128	4 h r	25
Pe 6 4	64	4 h r	20
Pe 3 2	32	4 h r	15
Pe 1 6	16	4 h r	10
L o n g 1 2 8	128	12 h r	27
L o n g 256	256	12hr	28
De b u g _ m d	128	10 m i n	29
De b u g sm	32	30 m i n	23

- **NQS Control Script (PERL 5)**
  - **Reads configuration file**
    - ♦ **Contains alternate queue configurations**
    - ♦ **Configuration selection based on time, day of week**
    - ♦ **Which queues are "on", "off", "backfill", etc.**
    - ♦ **Specifies global, complex and queue limits**
 - **Gathers system state: parses output of ps, grmview, qstat, psview**

- **NQS Control Script (cont'd)**
  - **Modifies NQS (via qmgr) to conform with selected configuration**
  - **Uses checkpoint/restart to switch between configurations**
    - ♦ **Up to 5 checkpoints done in parallel - average checkpoint of full machine in under 4 minutes!**
  - **Logs system state and all actions to time-stamped log file**

- **System Wide Checkpoint/Restart has many benefits**
  - **Facilitates queue scheduling shifts**
  - **Reduces lost time and lost work**
  - **Improves individual programs efficiency**
    - ♦ **One application on NERSC**
      - ➢ **Pre-processing - 30 minutes set up**
      - ➢ **Post-processing - ~30 minutes for clean-up**
    - ♦ **Changing from a 2 hour queue limit to a 12 hour limit**
      - ➢ **2 hour queue provided 6 hours of science computation**
      - ➢ **12 hour queue provided 11 hours of science computation**
    - ♦ **C/R allows multiple applications to get a share of time while allowing application to "perceive" a long run time.**
    - ♦ **83% more science work done for the same amount of time**

MPP Charging and Usage
FY 98-99
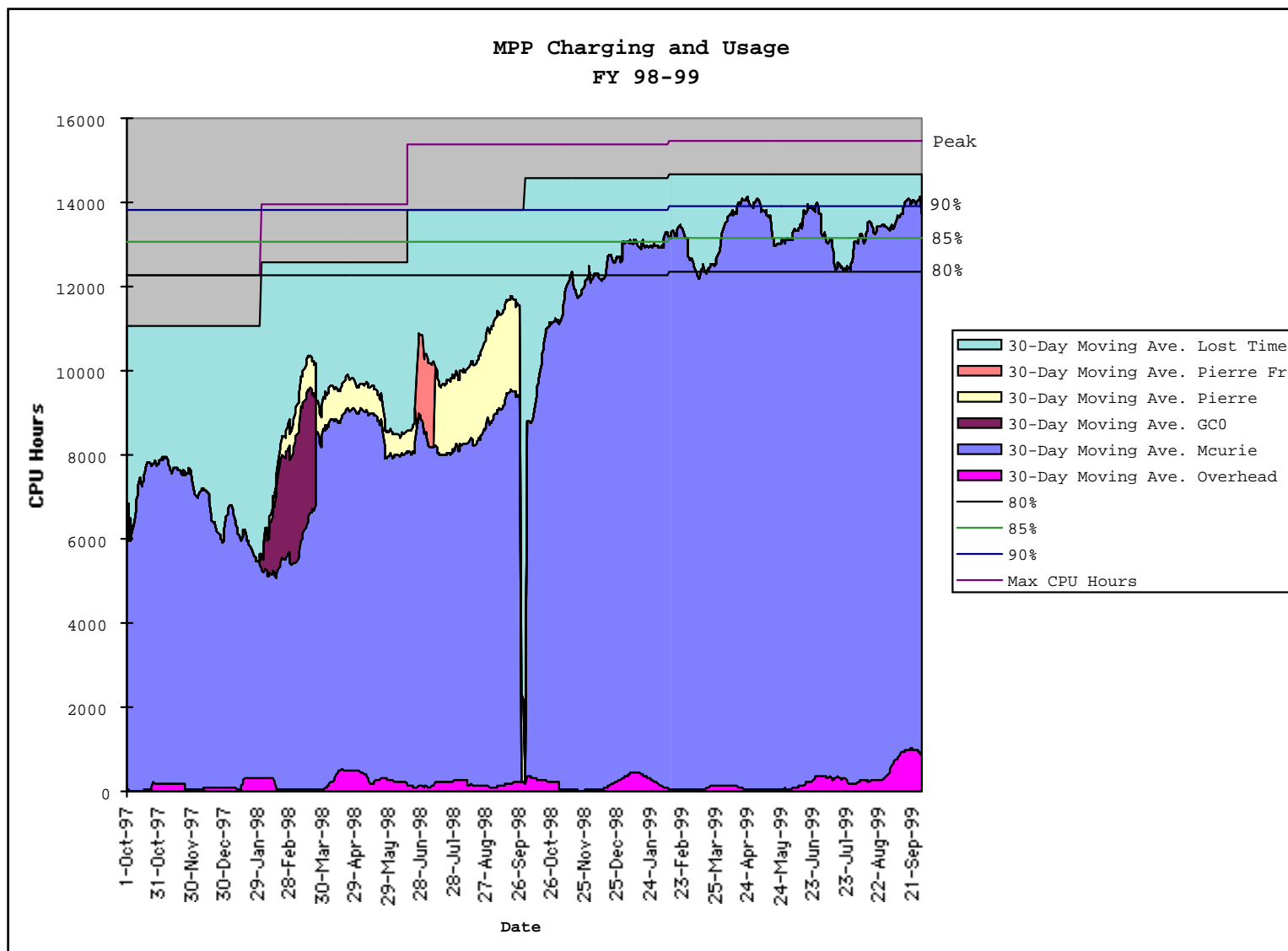
- **Vendor's Psched has become very stable**
- **Mods to GRM Service Limits are an effective means of managing the interactive workload**
- **Prime job feature is an effective tool critical for**
  - **Providing quick interactive response**
  - **Scheduling large jobs**
- **System management is simplified**
- **System-wide checkpoint/restart was integral to reducing lost time and lost work thus improving utilization**
- **Utilization is high (peak 92 - 97%). Nearly three years of joint effort**
- **Successfully completed paradigm Shift to massively parallel production processing**

- **Hardware**
  - **MPP: Homogeneous nodes**
  - **COTS: Heterogeneous slices of homogeneous nodes**
- **System**
  - **MPP: Single system image**
  - **COTS: Multiple identical systems**
- **Network Interconnect**
  - **MPP: Fast, proprietary**
  - **COTS: Slow, commercial**

- **Filesystem**
  - **MPP: Global**
  - **COTS: Shared and local**
- **N-Way Jobs**
  - **MPP: N-way job requires N CPUs**
  - **COTS: 1 node down does not stop N-way job**
- **Space, Cooling, Power Requirements**
  - **MPP: Densely Packed - Less space, more power, more cooling**
  - **COTS: Loosely Packed - More space, less power, less cooling**

- **Current**
  - **CPU: 1540 SPECint95**
  - **DISK: 4.2 TB**
  - **NET: 100 Mbs**
- **2 Years**
  - **CPU: 6000 SPECint95**
  - **DISK: 16 TB**
  - **NET: 1000 Mbs**
- **4 Years**
  - **CPU: ~15000 SPECint95**
  - **DISK: ~50 TB**
  - **NET: 1000+ Mbs**

- **NFS Mounted System - Eases Homogeneous System Configuration**

- **Mosix -- Kernal-level adaptive load-balancing and memory ushering**

- **Heterogeneous Slices Mapped Onto Heterogeneous Usage Patterns: e.g. Interactive, Short, Long Jobs**

- **Hardware Evolution**
  - **IBM RS/6000 SP: 304 SMP Nodes (2 cpus/node) $\Rightarrow$ IBM RS/6000 SP: 152 SMP Nodes (16 cpus/node)**
  - **Each node is a separate computer!!**
- **Storage Evolution**
  - **10 TB formatted global filesystem $\Rightarrow$ 20 TB formatted global filesystem**
- **Storage Usage**
  - **5 x 1.1 TB user home filesystems, 3.3 TB scratch space, 1.1 TB common software fiflesystem + extra space**
  - **How to manage 20 TB formatted disk space???**

- **Massively parallel tightly coupled system**
- **Distributed system management**
- **Managing for high availability**
- **Distributed user management**
- **Node maintenance (upgrades, installs, synchronization, configuration)**
- **Complex networking**
- **System-wide checkpoint/restart**

- ■ **Security**
  - ● **UNICOS and UNICOS/mk have Integrated Security Capabilities**
    - ♦ **User Database (udb)**
      - ➢ **Resource Limits**
      - ➢ **Password Management**
      - ➢ **Privileges**
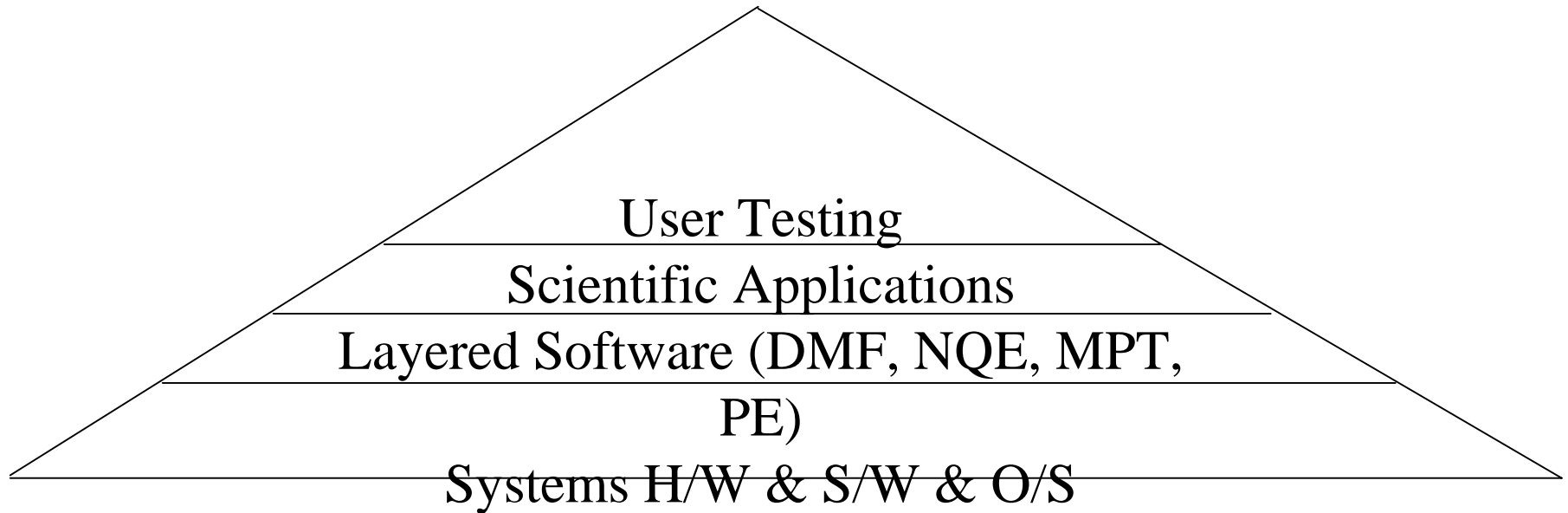   - ♦ **Security Log**

■ **Security (cont'd)**

● **User Accounts**

♦ **Disable for Non-Use**

♦ **Enforce Password Policy (8 characters, Special character, number)**

♦ **Educate Users**

♦ **Tell users to contact you if there has been any suspicious logins**

- **Security (cont'd)**
  - **Monitoring Tools**
    - ♦ **Tripwire**
    - ♦ **UNICOS MLS Security Logging**
    - ♦ **Spflick File Monitoring**
    - ♦ **BRO (packet sniffer)**
  - **Network Configuration Management**
    - ♦ **Know Your Network**
    - ♦ **Keep Current and Maintain a Network Diagram**

- **Security (cont'd)**
  - **TCP Wrappers**
  - **Install Firewalls Where They Make Sense**
    - **Staff Desktops**
    - **Special Purpose (Operator) Consoles**
  - **SSH**
    - **Local Mods Integrated with UNICOS 10.0 and UNICOS/mk 2.0.4**
      - **Over a year effort - unable to get vendor to port or support**
    - **Required for root access**
    - **Telnet will be turned off - December 1999**

- **Established and verified a Y2K baseline on each Computational System**

- **Setup Test Environment (alternate boot)**

- **Enormous amount of staff effort involved**

- **Additional system down time required**

- **Know your contingencies/options**

User Testing

Scientific Applications

Layered Software (DMF, NQE, MPT, PE)

Systems H/W & S/W & O/S

- **Balanced system means much more that just H/W. It means:**
  - **H/W that is sized right for your environment and application space**
    - **Disk space (swap, checkpoint, homes, /usr/tmp, system images)**
    - **CPU performance**
    - **Size of memory**

- S/W
  - ♦ O/S Stability
  - ♦ Current Compilers
  - ♦ Third Party Applications
  - ♦ Customizable Scheduling Scripts
  - ♦ Access to Source Code, Buildable Preferred
- Good staff (both vendor and NERSC) was critical to our success
- Nothing goes exactly as planned, especially the schedule. Need to know contingencies
- Single system image was a <u>big</u> positive attribute
- S/W performance improvements lags behind H/W performance improvements

- **As HPC sites move towards more clustering of SMP systems the probability of Tera-Scale computing on a "Super Computing Scale" is <u>finally</u> at hand. The challenge for HPC sites is to work together more and to work smarter with the vendor(s).**

- **We also need to retain our highly trained staff. Staffing may become our "Achilles' heel".**

■ **As commodity based H/W becomes more viable for Tera-Scale Super Computing the more complex/scalable the system S/W must become. From a system administrator perspective, a "turn-key" style of a Tera-Scale computer is not likely in the foreseeable near future. Much work is still at hand.**

We've all got our work cut out for us!